International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

CONTENT BASED VIDEO ACTIVITY CLASSIFIER

Aswany K H

[1] *P.G. Student, Department of Computer Science and Engineering, LBS Institute of Technology for Women, Trivandrum, Kerala, India.*

Abstract - *An important stream of research within computer vision is the ability to understand human activity, look and behaviour from a video. The human activity recognition has gained a variety of applications such as content base retrieval, human motion analysis, electronic video surveillance and visual enhancement. Selection of extracted features plays an important role in content-based video activity classifier. These features are intended for selecting, indexing and ranking according to the potential interest to the user. The major problem here is that how to analyse and search within videos based on their content, with minimum human intervention. The proposed method aims to develop such a system where user will be able to search based on the semantic content of the video like in a query response mechanism.*

***Key Words*: Computer vision, human motion analysis, query response mechanism, semantic content, video surveillance.**

1.INTRODUCTION

The goal of activity recognition is to detect human activities in a real-time or off line video. Human activity is highly complex and diverse because of which human activity recognition becomes challenging. As a first step, activity model must be built. Probability-based algorithms, Hidden Markov Model and the Conditional Random Field are among the most popular modeling techniques. The proposed system uses Hidden Markov model for human activity recognition.

HMM are not algorithms. But the question whether HMM belong to "clustering" or "classification" family is as pointless as asking whether the Gaussian distribution is supervised or unsupervised learning. While using HMM, we will be able to know good parameter estimation and marginal distribution computation algorithms since it is a specific kind of probability distribution over sequences of vectors. In Bayesian framework, HMM can be used for classification and being model with hidden states, clustering of the training data can be recovered from their parameters. More precisely, HMM can be used for both classification and clustering. For example, you have a large database of digits ("one", "two", etc.) and need to build a system capable of classifying an unknown input. For each class in the training data, you will build a model and will end up with 10 different models. Then, to perform recognition, you compute the 10 likelihood scores, and the digit will be matched to the model with the highest score.

HMM can be used in an unsupervised fashion also. For that, input a sequence and train a k state HMM on it. To get the most likely class associated with each input vector, run the Viterbi algorithm during the training process. Then your input sequences will be clustered into k classes. For example, in the case of video sequences, color histogram of each frame will be extracted and at the end, video will be breakdown into homogeneous temporal segments corresponding to scenes. This technique is commonly used in unsupervised analysis of video. The proposed system also analyses video in the unsupervised fashion.

2. RELATED WORK

Regarding the human activity recognition, we must distinguish between supervised and unsupervised classification methods. In supervised approach, classification is done based on the labeled training data and its use has gained favorable results [1]. The Gaussian Mixture Models (GMM) approach [3], k-Means algorithm [2] and the one based on Hidden Markov Model (HMM) [4], [5] or HMM with GMM emission probabilities [6] can be used in unsupervised approach. Both the GMM and the HMM approaches use the EM algorithm [7]. In the proposed system, the unsupervised approach for human activity recognition is used. More precisely, the proposed approach is based on a Hidden Markov Model. The most likely hood occurrence of each activity is calculated using the Viterbi algorithm [8]. Selection of intended features plays an important role in activity recognition. Mean, median, standard deviation and correlation are most popularly considered features in activity recognition [11]. As a first step of feature extraction dimensionality reduction must be done and most widely adopted is Principal Component Analysis [12].

This paper is organized as follows; section 3 presents the design considerations and description of the proposed algorithm. Section 4 is the implementation details of the proposed system. In section 5, the performance of the proposed system is evaluated and compared to some of the popular supervised and unsupervised techniques of human activity recognition.

3.PROPOSED ALGORITHM

3.1 Design Considerations

- Considered only 3 activities at the beginning.
- HMM in the unsupervised fashion is considered.

- Crowded environments are not considered.
- Keeping track of all the statistical features.
- Videos with total frame less than two is ignored.

3.2 Description of the Proposed Algorithm

Aim of the proposed method is to develop a system where user will be able to search based on the semantic content of the video like in a query response mechanism. These are the software prerequisites of the proposed algorithm.

- Python >= 2.6
- NumPy >= 1.9.3
- SciPy >= 0.16.0
- scikit-learn >= 0.16
- Matplotlib >= 1.1.1
- Pytest >= 2.6.0
- hmmlearn
- opencv

Step 1: Training Phase: The dataset required for the training is selected and for each training video, frame by frame segmentation is done. Activities under considerations are running, walking and hand clapping. The system is trained with the KTH action database and for every activity and HMM model is built.

Step 2: Feature Extraction: As a first step in feature extraction, dimensionality reduction is done since the features associated with multimedia data like video is huge. So, to decrease the complexity, it must be reduced to smaller dimensions. Through feature extraction, model parameters are estimated. For recognizing each activity its position, motion intensity and action descriptors are considered as observed parameters. With these observed parameters, a sequence of hidden states is derived by using Viterbi algorithm. With the help of the forward-backward algorithm, a set of probabilities for each frame is generated. Finally mean value of these probabilities are calculated and is assigned to training video.

Step 3: Testing Phase: When a test case is inputted, its likelihood probability is generated and from these insight, best matched model is invoked and will thereby predict the activity within that video. Along with the predicted class both actual class and test accuracy are also generated.

4. IMPLEMENTATION DETAILS

Step 1: Import all the required libraries.
Step 2: Define the allowed classes.
allowedclass=["running","walking",
"handclapping"]
Step 3: Get the required data set.
→Universal data set KTH Action database is used.
Step 4: Start the training phase.
→Video frames are extracted and stored in sequences.txt
→For each class in the allowed class,
→Initialize class function, class length and model.
→Model used is: Hidden Markov Model with Gaussian emissions.
→For each line in the sequences.txt
→Using split (), larger string is broken into smaller ones and stored in tok.
→If length of tok>0,
→function name is generated and stored.
→For each frame, features are extracted and model is built.
Step 5: Feature extraction and model construction.
→Dimensionality reduction is done.
→Load each video and extract frame and frame count.
→For each image, best choice is found out with model parameters.
→Using bgsegm in opencv, background subtraction is done.
→Using various inbuilt functions in opencv, feature are extracted and returned to main function.
Step 6: For each frame likelihood probability of activities is generated.
Step 7: For each class in the allowed classes, probability of class is calculated as mean probability of all their frames.
probClass[c] = np.mean(prob[c])
Step 8: In test case, when a new video is given, it is matched with the corresponding model and its test accuracy, predicted class and actual class is displayed.
Step 9: Test accuracy= test_log_correct*100.0 /test_log_total
→test_log_correct and test_log total are initialized to zero.
→Every time when a probability is generated for each frame, test_log_total is incremented by one.
→If actual class = predicted class, then test_log_correct is incremented by one.

5. RESULTS AND DISCUSSIONS

5.1 Performance Evaluation

This section presents the performance evaluation of content based video activity classifier. The proposed algorithm is applied to a wide range of videos from KTH action database. Figure 1 represents some parts of the segmentation results of a test case. Actual content of the given test video is running, and these are some of the frames after segmentation.

The allowed classes in the proposed method are running, walking and hand clapping. For estimating test accuracy 100 video set are taken. For evaluating parameters such as accuracy, misclassification rate, precision, etc., a confusion matrix is generated. There are two possible predicted classes: "YES" and "NO". "YES" would mean actual content of the video is within the allowed class and "NO" would mean actual content of the video is other than that of the allowed class. The classifier made a total of 100 predictions. That is, 100 videos are being tested for recognizing human activity.

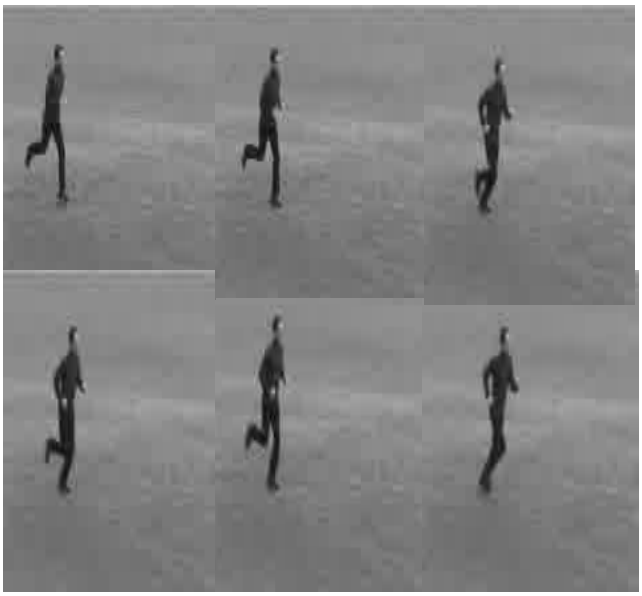


Fig-1: Segmentation result of a test case video.

Out of those 100 cases, the classifier predicted "yes" 52 times, and "no" 48 times. In reality, 50 test videos are not within the allowed class and another 50 set of videos are within the allowed class. TP represents the true positive which is the cases where the actual activity of the video is within the allowed class, and they are predicted correctly also. TN represents the true negative which is the cases where the actual activity of the video is not within the allowed class and they are also predicted correctly. FP represents the false positive which is the cases where the actual activity of the video is not within the allowed class but it is predicted as running, walking or hand clapping. FN represents the false negative which is the cases where the actual activity of the video is within the allowed class, but it is predicted wrongly.

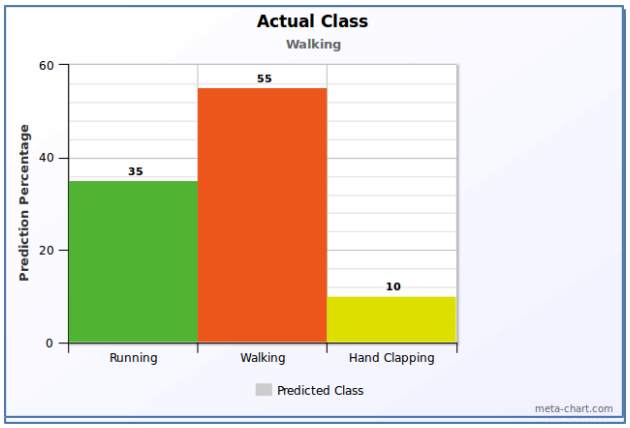
- Accuracy: (TP+TN)/total = (49+47)/100 = 0.96
- Misclassification Rate: (FP+FN)/total = (3+1)/100 = 0.04
- Sensitivity: TP/actual yes = 49/50 =0.98
- False Positive Rate: FP/actual no = 3/50 = 0.06
- Specificity: TN/actual no = 47/50 = 0.94
- Precision: TP/predicted yes = 49/52= 0.94
- Prevalence: actual yes/total = 50/100 = 0.5

Table -1: Confusion Matrix

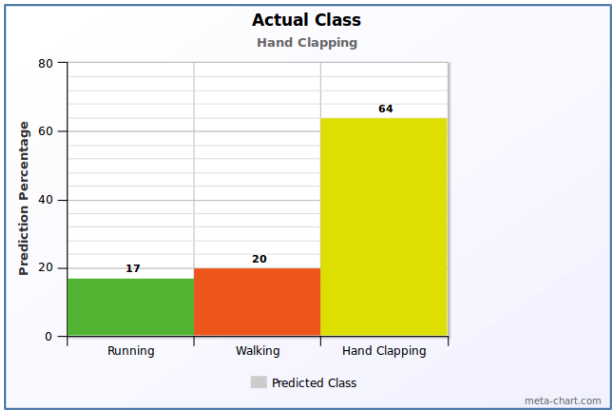
n=100	Predicted: NO	Predicted: Yes	
Actual: NO	TN=47	FP= 3	50
Actual: YES	FN=1	TP=49	50
	48	52	

The proposed system can recognize human activities such as running, walking and hand clapping. i.e. the actual class always matches with the predicted class with a considerable amount of probability. But even though the highest probability is assigned to the actual content of the video, traces of other classes are also detected. For each allowed class, this distribution is shown in the histogram.

(a)



(b)



(c)

Chart -1: Histogram representation

Semantic content of the videos is labelled as actual class. But it shows considerable amount of probability to other classes

also. (a) Actual class is running. Highest probability is assigned to running. (b) Actual class is walking. Highest probability is assigned to walking. (c) Actual class is Hand Clapping. Highest probability is assigned to hand clapping.

5.2 Comparison of Machine learning techniques for Human Activity classification

The latest developments in multimedia generate numerous amount of data every second. If this data is analyzed efficiently, it can disclose a lot of insights. To analyze multimedia data such as video, tremendous machine learning techniques are available.

(1). Decision Tree
For Human action recognition, one of the commonly used algorithm is decision tree. Accelerometer embedded in the smart phones are used for classification in most of the work. However, if the new observation doesn't match with the training observation, the accuracy will suffer [14].

(2). Support Vector Machine
SVM is a supervised machine learning algorithm. SVM classifier separates the events into different classes based on the target output with maximum marginal hyperplane. But in SVM the non-linear inputs may affect the precision [13].

(3). K-Means
K-means is a popular clustering technique in machine learning which will cluster n objects into k partitions where k<n [9]. Value of k is a pre-chosen one and there is no method for finding the exact value of k. Since it is an iterative process it may affect the result.

(4). Fuzzy C-Means
This is another clustering technique where with certain membership value all the data points in the dataset belongs to every cluster [9]. Initially cluster centers are guessed and iteratively mean value of clusters is marked. The iteration aims at the minimization of objective function which calculates the distance between cluster center and given data point. Calculating the degree of association is a tedious task here.

(5). Convolutional Neural Network
Human activity can be recognized based on convolutional neural network which make use of both local dependency and scale invariance of the signal [10]. In the case of video, the relationship between the nearby pixels of a frame in its base level is considered as local dependency. In activity such as walking a person may walk with different motion intensity that is with different scale which is treated as scale invariance. With different datasets, it shows different accuracy.

Table -2: Comparison of performance of 5 algorithms which are commonly used for Human Activity Recognition.

Sr	Algorithm Used	Accuracy
1	Decision Tree	86.86
2	Support Vector Machine	96.4
3	K-Means	80
4	Fuzzy C-Means	78
5	Convolutional Neural Network	97.1

6. CONCLUSIONS

Video surveillance has gained a lot of importance today. But analyzing the whole video manually is a time-consuming process. In this paper, a system is proposed where user can search within the video based on its content with minimum human intervention like in a query response mechanism. As an initial step the system is constrained in human activity recognition only. HMM is used to detect human activity within the video. The proposed system was trained to detect human activity and this was carried out for three different activities such as running, walking and hand clapping. The test accuracy and experimental results shows that HMM is well suited for Human activity recognition. Its future scope includes, building a system which consists of a mechanism to identify videos with obscene content on-line, full length feature film that have been illegally uploaded, crime occurrence, property loss by mistake etc. within a video.

REFERENCES

[1] Altun, K., Barshan, B., Tuncel, O., "Comparative study on classifying human activities with miniature inertial and magnetic sensors", Pattern Recognition, 43(10),3605-3620, 2010.

[2] Duda, R.O. Hart, P.E., and Stork, D.G., Pattern Classification (seconded). A Wiley Inter Science Publication, John Wiley& Sons, 2000.

[3] Allen, F.R., Ambikairajah, E., Lovell, N.H., and Celler, B.G., "Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models", Physiol.Meas., 27(10),935-951, 2006.

[4] Lin, J.F.S., and Kulic, D., "Automatic human motion segmentation and identification using feature guided hmm for physical rehabilitation exercises", In: Robotics for Neurology and Rehabilitation, Workshop at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2011.

[5] Rabiner, L.R., A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77(2), 257-286, 1989.

[6] Mannini, A., and Sabatini, A., "Machine learning methods for classifying human physical activity from on-body accelerometers" Sensors, 10, 1154-1175, 2010.

[7] Dempster, A.P., Laird, N.M., and Rubin, D.B., "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, B, 39(1), pp. 1-38, 1977.

[8] Viterbi, A.J., "Error bounds for convolutional codes and asymptotically optimum decoding algorithm", IEEE Transactions on Information Theory 13(2), 260-269,1967.

[9] Maleeha Kiran, Lai Weng Kin, Kyaw Kyaw Hitke Ali, "Clustering Techniques for Human Posture Recognition: K-Means, FOM and SOM", Recent advances in Signals and Systems, us/e library conferences, ISBN: 978-960-474-114-4, 2009.

[10] Zeng , M, ; Nguyen, L.T. ; Yu, B. ; Mengshoel, O.J. ; Zhu, J. ; Wu, P. ; Zhang, J. "Convolutional Neural Networks for human activity recognition using mobile sensors" In Proceedings of the 6th IEEE International Conference on Mobile Computing Applications and Services (MobiCASE), Austin, TX, USZ, 6-7 November 2014.

[11] D. Figo, P.C Diniz, D.R. Ferreira, and J.M. Cardoso. "Preprocessing techniques for context recognition from accelerometer data". Personal and Ubiquitous Computing, 14(7): 645-662, 2010.

[12] Z. He and L. Jin. "Activity recognition from acceleration data based on discrete cosine transform and SVM". In Systems, Man and Cybernetics. SMC 2009. IEEE International Conference on, pages 5041-5044. IEEE 2009.

[13] Sergios Theodoridis, Michael Mavroforakis, "Reduced Convex Hulls: A Geometric Approach to Support Vector Machines [Lecture notes]", Signal Processing Magazine IEEE, vol. 24, pp. 119-122, 2007, ISSN 1053-5888.

[14] Tatiana Jaworska, "Application of fuzzy rule-based classifier to CBIR in comparison with other classifiers" Fuzzy Systems and Knowledge Discovery (FSKD) 2014 11th International Conference on, pp. 119-124, 2014.